International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; 8(5): 109-115

https://www.mathsjournal.com
Received: 04-07-2023
Accepted: 09-08-2023

Pooja
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Vishal Mehta
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Sandeep Kumar
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

RR Kushwaha
Department of Agricultural Economics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Corresponding Author:
Vishal Mehta
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Regression models for forecasting pulses area and production in India

Pooja, Vishal Mehta, Sandeep Kumar and RR Kushwaha

Abstract
India is the largest producer as well as consumer of pulses. In spites of highest position of India in the production of pulses, it is importing 3-4 million tonnes (MT) of pulse every year to meet its domestic demand. Pulses are grown in an area of 22-23 million hectares with an annual production of 13-18 MT. This is about 124355 tonnes (32.72%) of the total production in the year of 2012- 13 (Ministry of Agriculture, GOI, 2013-14). Estimating crop production and predicting crop area are the key ways to take the necessary actions to preserve food security and solve environmental issues. Crop production forecasting is the skill of predicting the future values of produced crops before they are harvested. In this study, the main objective was to develop a clear portray about the area and production of pulses in India. The art of anticipating the future values of cultivated crop in advance before its harvest is called as forecasting of crop production. In this study, the main objective was to develop a clear portray about the area and production of pulses in India. The state with largest area under Pulses cultivation is Rajasthan (6.34 million hectares), Madhya Pradesh (4.76 million hectares), Maharashtra (4.19 million hectare), Karnataka (3.11 million hectares) and Uttar Pradesh (2.37 million hectares). Regression modeling was applied for both area and production of Pulses in India for a period 71 years (1951-2021) using linear, logarithmic, quadratic, cubic, power and exponential models. The results concluded that cubic model shown better performance compared to other models for prediction of area and production of pulses for India with maximum accuracy.

Keywords: Pulses, area, production, regression modeling, RMSE (Root mean square error) and MAPE (mean absolute percentages error)

Introduction
In India more than a dozen of pulse crops including chickpea (40%) pigeon pea (18%), urd bean (11%), mung bean (9%), lentil (8%), field pea (5%) and others are grown on 22 - 24 million hectares producing 13 - 15 million tonnes of grain with an average productivity of 600 - 650 kg/ha. In India more than a dozen of pulse crops including chickpea (40%) pigeon pea (18%), urd bean (11%), mung bean (9%), lentil (8%), field pea (5%) and others are grown on 22 - 24 million hectares producing 13 - 15 million tonnes of grain with an average productivity of 600 - 650 kg/ha. The largest percentage of India's output of pulses is contributed by the state of Madhya Pradesh. Which ones are the healthiest? In comparison to other protein sources, pulses are far more inexpensive and a substantial source of protein. The family of legumes contains more than 13,000 species and more than 600 genera [see, https://pulses.dac.gov.in/areaproduction.aspx]

All types of dried beans, such as kidney beans,lima beans, butter beans, and broad beans, are common pulses. Pulses also include all types of lentils including chick peas, cowpeas, black-eyed peas, and pigeon peas. Pulses are used in staple meals and cuisines from around the world, including Indian dal, which uses chickpeas, a classic full English breakfast, and hummus in the Mediterranean, which uses chickpeas (peas or lentils). The majority of pulses crops come from India. India is the leading producer of Pulses contributing more than 25 percent of total Pulses produced around the world.

All pulses are nutrient-dense, offering a substantial amount of fiber, protein, minerals (suchas iron), and vitamins (such as folate), despite the fact that the nutritional value of different species may vary slightly. In addition, unlike their oilseed legume relatives, pulses are low in fat and free of gluten and cholesterol.

Fiber: When compared to other carbohydrate-rich meals like rice, white bread, or potatoes, pulses have a lower glycemic index while having a high concentration of carbohydrates (50 to 65%). This is because of the fiber in pulses, which also leads in delayed digestion. Whole-grain cereal crops like brown rice and oats are often promoted for their fiber content. However, pulses contain 2-3 times more fiber and protein than cereal crops.

Protein: Pulses are one of the richest sources of plant-based protein since they contain roughly 1 gramme of protein for every gramme of fiber. Minerals: Iron, potassium, magnesium, and zinc are a few of the important minerals included in pulses. Vitamins: Pulses are also rich in B vitamins, such as folate, thiamin, and niacin.

Table 1: Nutritive Value of Pulses

Nutritional evaluation of various pulsus in a 100 g.				
Nutrient	**Chickpeas**	**Common Beans**	**Dry peas**	**Lentil.**
Energy (kcal)	120	120	116	104
Protein (g)	6	8	8	8
Fat (g)	2	1	0.5	0
Carbohydrate (g)	20	21	21	18
Fiber, total dietry(g)	4	6	8	7
Iron (mg)	1.4	2	1.3	3

Regression is a key tool in data science for forecasting a variety of continuous targets, helping to ensure that this task is not overstressed and is not perceived in the real world as being all-encompassing. Even nowadays, the studies linked to regression methods are rigorously evaluated by academics since it gains maximum attention in important.

In the present study, Regression modelling was used to fit and forecast the area and production of pulses. Regression model use for area and production forecasting was the study's main goal. The main outcome expected from the current study was to obtain complete sketch about future anticipation on pulses which may help the farmer to make and adopt the plans and also for the government for making related policies.

Review of literature

Supriya *et al.* (2023) [11] constituted an essential component of vegetarians' protein rich diets in India. Yearly data were collected from Agriculture Statistics at a glance, 1970 to 2019 were used for forecasting up to 2029. India was found suitable to forecast the future of lentil with an 80% and 95% accuracy level and according to the analysis of instability, the instability is increasing in all three states and India, but the instability was decreasing in Uttar Pradesh.

Yadav *et al.* (2023) [7] revealed that develop a clear portray about the area and production of nutria-cereals in India. India is the leading producer of nutria-cereals which is almost 33.3 percent of total nutria cereals produced all over the world. The largest nutria cereals-producing state in India is Rajasthan which almost contributes 41.03 percent of total nutria cereals produced with in the country. Regression modeling was applied for both area and production of nutri cereals in India for a period 70 years (1951-2020) using linear, logarithmic, quadratic, cubic, power and exponential models. The results concluded that cubic model shown better performance compared to other models for prediction of area and production of nutria cereals for India with maximum accuracy.

Yadav *et al.* (2022) [6] estimated that the study employed modeling method (Box-Jenkin's ARIMA model) to anticipate nutria cereals area in India. The optimal Arima arrangement was shown to be (0, 1, 1). Furthermore, by applying ARIMA (0, 1, 1) model to our time series data, we attempted to anticipate the future nutria cereals Area as accurately as possible for a period up to ten years nutria cereals Area according to the anticipated results. It will continue to fall at a peace of about 1% each year on a average.

Sah *et al.* (2021) [9] reported that the temporal changes in the area, production and productivity of major pulse crops in Bundelkhand region of UP during last two decades (2000-2020). Urdbean registered highest decline (-6.52%) in productivity during the raid period. In contrast, moongbeam crop followed by urdbean recorded highest growth in area (3.48%, 3.02%), and production (3.61%, 2.71%), while field pea registered highest growth in productivity (1.96%). Instability of pulse production over the study period was high (30.31%), however the instability in pulse area was low (11.7%). Analysis of inter district variation in pulse performance reflected decline in area and production of pulses in all districts except in Lalitpur. Inter district variations helped in identifying the potential districts in the Bundelkhand region of UP state.

Mandal *et al.* (2020) [4] conclude that the current state of pulse production in general and enterprise in West Bengal and estimate the nature, level and direction of interactive relationships among the dependent variables, wages and man-days required in pulse crop cultivation, a set of economic and ecological variables were selected for the study.

Vani and Mishra (2019) [12] evaluate that For Indians, pulses are a source of protein, but supply has not kept up with the rising demand of the country. Among efforts to enhance pulses production, role of irrigation as a critical input has not been given due importance. It was found that yield and area not-irrigated effect accounts for 52 per cent of growth of pulses which is not suitable for sustainability of pulses production system. Area not-irrigated effect contribution was 13.69 percent on pulses production.

Kumar and Singh (2016) [3] find the result, the world's biggest producer, consumer, and importer of pulses is India. In India, pulses are grown in around 23.26 million hectares of area and producing 18.34 million tonnes of pulses annually. After analysis of production of pulses by structural time series model using SAS, pulses production forecast for the year 2020 to be near about 19.38 million tones with upper and lower confidence limit 16.83 and 21.94 million tonnes respectively and it shows that there is an increasing trend for production of pulses in India.

Narayan and Kumar (2015) [5] conclude that, a diet with a balance of nutrients should include pulses. These are the principal source of protein for vegetarians. India is the world's largest producer of pulses, followed by Canada. Brazil produces large beans only. Pulses are the second main source of protein after cereals in Indian diet. This paper analyses the status of pulses growth, and constraints of technology inadequacy as well as policy reform.

Sharma *et al.* (2013) [10] reported that this study uses time series data on area, production, yield, and trade for the years 1980-1981 to 2008-2009 to analyze the rise and trend of pulse production in India. The period has been divided into two parts *viz.*, before (1980-81 to 1994-95) and after (1995-96 to 2008-09). Linear, semi-log growth function and compound growth function were tried to examine the growth and trend. Area of pulses has been most stable with few exceptions as compared to production and yield of pulse corps indicating thereby technological interaction in the pulse production. On other hand, instability in export and import has, however, been quite high during the period of study.

Reddy and Reddy (2010) [8] revealed that with a yield of roughly 600 kg/ha, the yearly production of pulses in India ranges from 11 Mt to 15 Mt. Due to the wide gap between supply and demand, import of pulses has increased from 0.38 Mt in 1993 to 2.82 Mt in 2008. Lentil is an important rabi pulse crop with a production of 0.85-0.95 Mt in India, after gram.

Data and methodology

Collection of data

The secondary data required for the study pertaining to area, production and yield of pulses in India of 71 years (1951-2021) was collected from the Agricultural Statistics at Glance 2021, Government of India (GOI).

Source of Data: Directorate of Economics & Statistics, DAC&FW, New Delhi. https://eands.dacnet.nic.in/

Regression Analysis

In regression analysis, relationship associated between response variable and explanatory variables are obtained after undergoing analysis of the data. In this study, the variables selected for undergoing analysis are year, area and production. Area and production of sorghum are considered as response variables and year is recognized as explanatory variable for this study.

The different steps undertaken for conducting this study is illustrated as a flow chart in Fig. 2. The collection of data for area and production of pulses was done for a period of 71 years (1951-2021) from Agricultural Statistics at a Glance 2021, Government of India. The modeling of area and production of pulses was undergone by using simple regression analysis. The future observation of response variable was predicted using the relationship existed in among an independent and a dependent variable. Therefore, the analysis was conducted with an assumption that the future values predicted for area and production of pulses would follow the past trend. The easiness and ability to predict any data accessible was told as the main asset of this approach.

Flowchart of solving Regression analysis

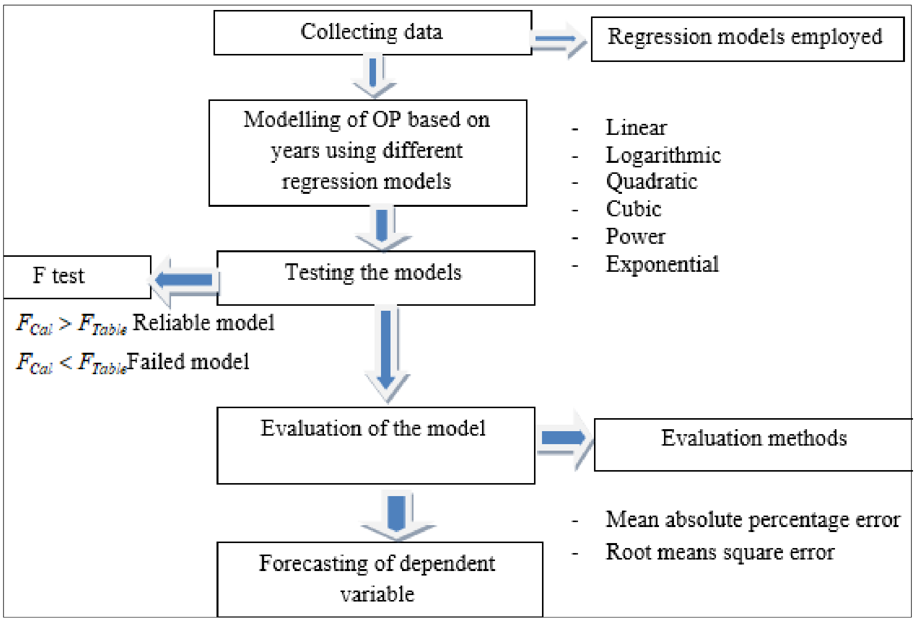

Fig 1: Flow chart of the study

For this study, regression analysis methods including both linear and non-linear methods like linear, logarithmic, power, exponential, quadratic and cubic regressions were utilized for fitting suitable models for area and production. It is noted that straight-line relationships exist for a linear model whereas it is curvilinear for a non-linear model. The data was split into two different datasets, training and testing datasets. The training dataset includes data for a period of 66 years (1951-2016) which was used for fitting the model whereas testing dataset consists of data for a period of last 5 years (2017-2021) used for evaluation of fitted regression model.

The regression models used in this study including both linear and non-linear models are mathematically expressed as:

Linear regression model:

$$Y_t = a + bt + \varepsilon_t \quad (1)$$

Quadratic regression model:

$$Y_t = a + bt + ct^2 + \varepsilon_t \quad (2)$$

Cubic regression model:

$$Y_t = a + bt + ct^2 + dt^3 + \varepsilon_t \quad (3)$$

Logarithmic regression model:

$$Y_t = a + b\ In(t) + \varepsilon_t \quad (4)$$

Power regression model:

$$Y_t = at^b + \varepsilon_t \quad (5)$$

Exponential regression model:

$$Y_t = ae^{bt} + \varepsilon_t \quad (6)$$

Where:

Y_t indicates the dependent variable *i.e.*, area or production, t denotes the independent variable, Time in years, a is the intercept, $b, c,$ and d are representing the regression coefficients and ε_t is the error term, $\varepsilon_t \sim \mathrm{N}(0, \sigma_e^2)$.

The verification of fitted model is completed by determining coefficient of determination (R^2) and an $F-Test$. R^2 was used to obtain total changes in a response variable due to regression model. The efficacy of model fitted was indicated by a higher (R^2 value. In order to undergo acceptance of fitted model and independent variable sophisticated in the model, F test was undergone. $F-Test$ was used to identify whether fitted model was significant or not. If the *F*-value obtained after undergoing *F* test is higher than the *F* table value (from *F* distribution table), the model is accepted or otherwise rejected.

The best forecasting model was identified by checking the work of regression models established for the data. Least value for RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentages Error) was the selected criteria for checking the work of fitted models [Eq. (7) and (8)]. Smaller values of RMSE and MAPE indicate the better prediction.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}{n}} \tag{7}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_t - \bar{Y}_t}{Y_t}\right| \tag{8}$$

Where Y_t is the observed value and $\hat{Y}_t$ is the predicted value, n is the total number of measurements.

Table 2: Linear and Non-linear models of Pulses Area in India

Area Models	Parameter				Criteria			
	a	b	c	d	R^2	$Adj.\ R^2$	RMSE	MAPE
Linear	-75.15*	0.04956*	---	----	0.240*	0.229*	1.806	0.0593
Logarithmic	-721.96*	98.13873*	---	---	0.238*	0.2278*	1.807	0.0057
Quadratic	6694.8528*	-6.77	0.0017	---	0.337*	0.3176*	1.687	0.057
Cubic	-1339265*	2026.66*	-1.02*	0.00*	0.650*	0.635*	1.223	0.0027
Power	-27.42*	4.02488*	---	---	0.234*	0.223*	0.073	0.0024
Exponential	-0.89155*	0.002032*	---	---	0.235*	0.224*	0.075	0.0025

* Indicate significant at 5% level of probability.

Table 3: Linear and Non-linear models of Pulses Production in India

Production Model	Parameter				Criteria			
	a	b	c	d	R^2	$Adj.\ R^2$	RMSE	MAPE
Linear	-263.926*	0.139*	---	----	0.593*	0.5877*	2.3667	0.1288
Logarithmic	-263.926*	0.139*	---	----	0.593*	0.5877*	2.3667	5.6011
Quadratic	15656.69*	-15.895	0.004	----	0.760*	0.7533*	2.3666	5.6011
Cubic	-1228815*	1864.21*	-0.94*	0.0002*	0.844*	0.8369*	1.466	2.1508
Power	-142.72*	19.1304*	---	----	0.642*	0.63687*	2.2569	0.1158
Exponential	-16.6144*	0.009651*	---	----	0.644*	0.63929*	2.2475	0.1155

* Indicate significant at 5% level of probability.

Table 4: Results of the Validation Tests for the Developed Models

Validation test	Regression type					
	Linear	Logarithmic	Quadratic	Cubic	Power	Exponential
t (area)	435.64	425.642	954.349	1018.67	290.60	297.242
t (production)	12.241	11.9811	93.597	61.486	16.952	17.293

Modeling of pulses area and production

The results of the regression analysis using different methods are described in Table 2. The R^2 values lie above 0.969 and 0.870 for fitted regression models, suggested an immense degree of ling among area and years, production and year respectively. The determination coefficients also revealed that only limited percent (0.011–0.010%) of the variation in the area and production are due other unknown factors rather than the predictors (see Eqs. 1 to 6). According to model evaluation criterion, cubic model is best fit for forecasting for area and production of sorghum (Eqs. 9 and 10).

Forecasted Cubic model is Area

$$\hat{Y}_{Area} = -1339256 + 2026.66t - 1.02t^2 + 0.01t^3 \tag{9}$$

$$\hat{Y}_{Production} = -1228815 + 1864.21t - 0.94t^2 + 0.01t^3 \tag{10}$$

Forecasted Values of Pulses Area and Production and Residuals Plots by Cubic model

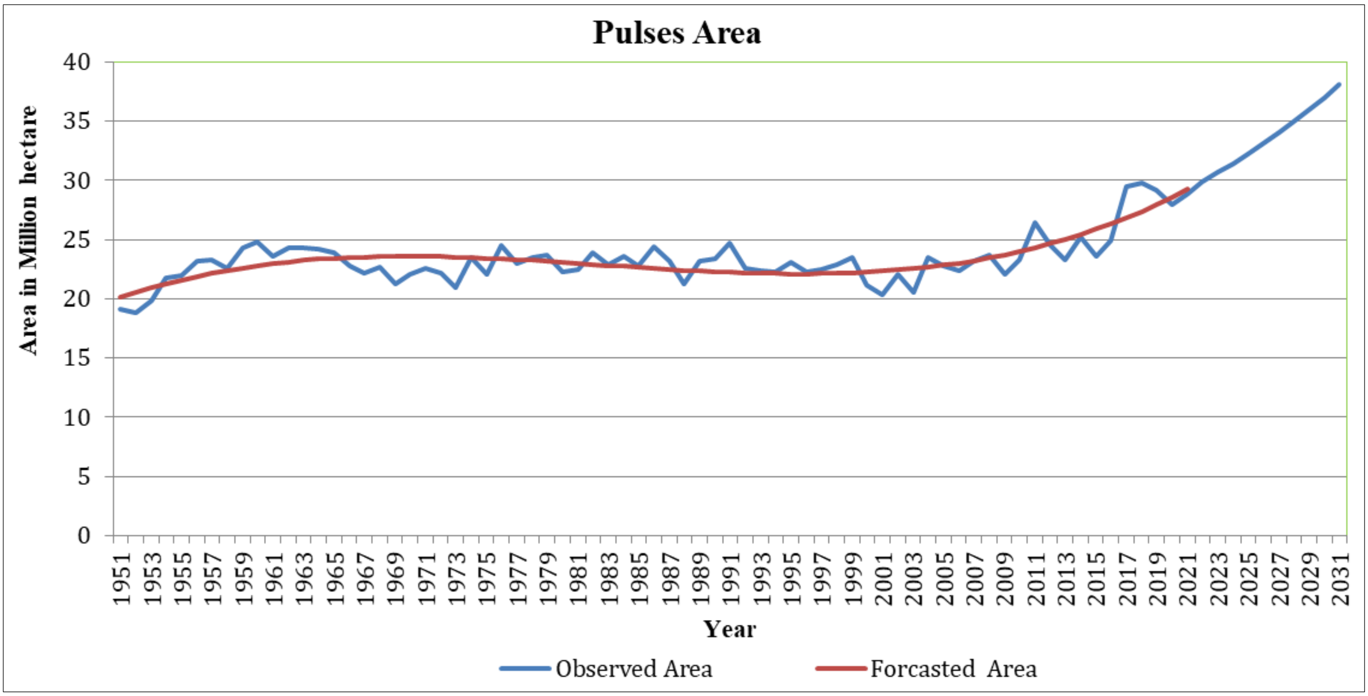

Fig 2: Actual and Forecasted Value of Pulses Area in India

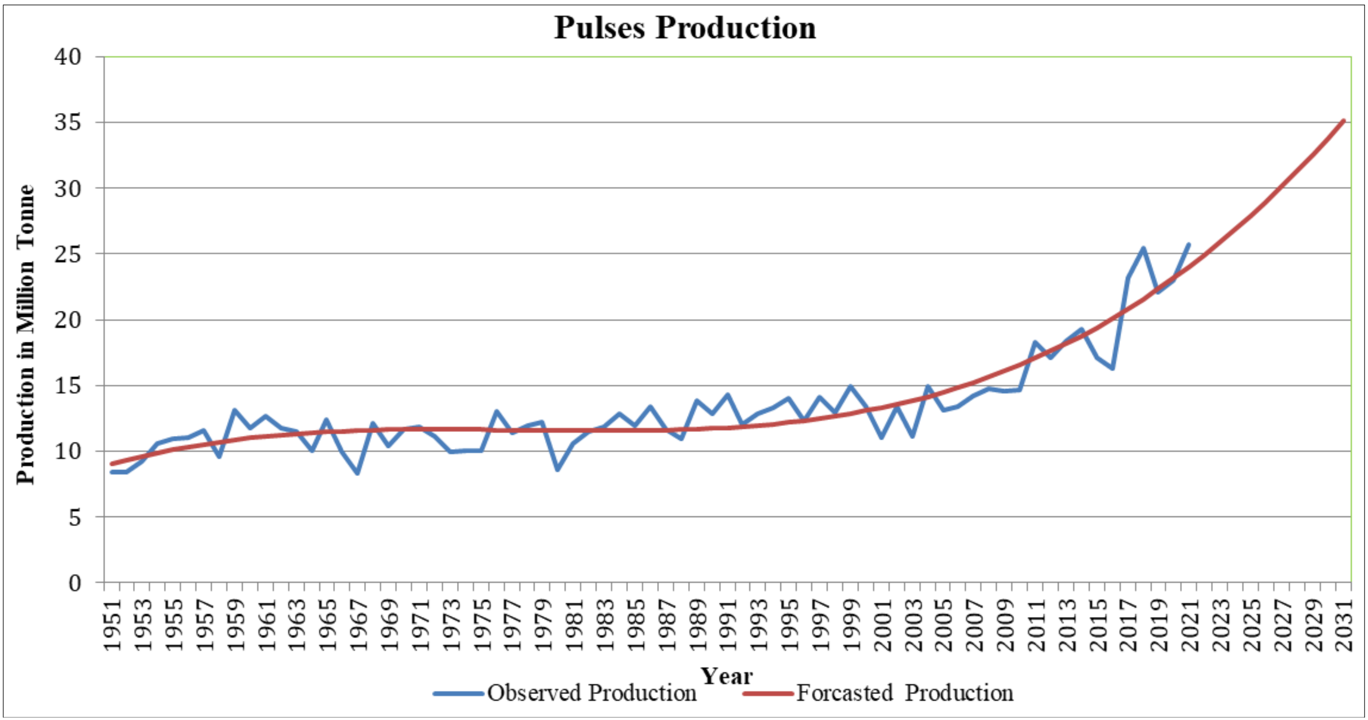

Fig 3: Actual and Forecasted Value of Pulses Production in India

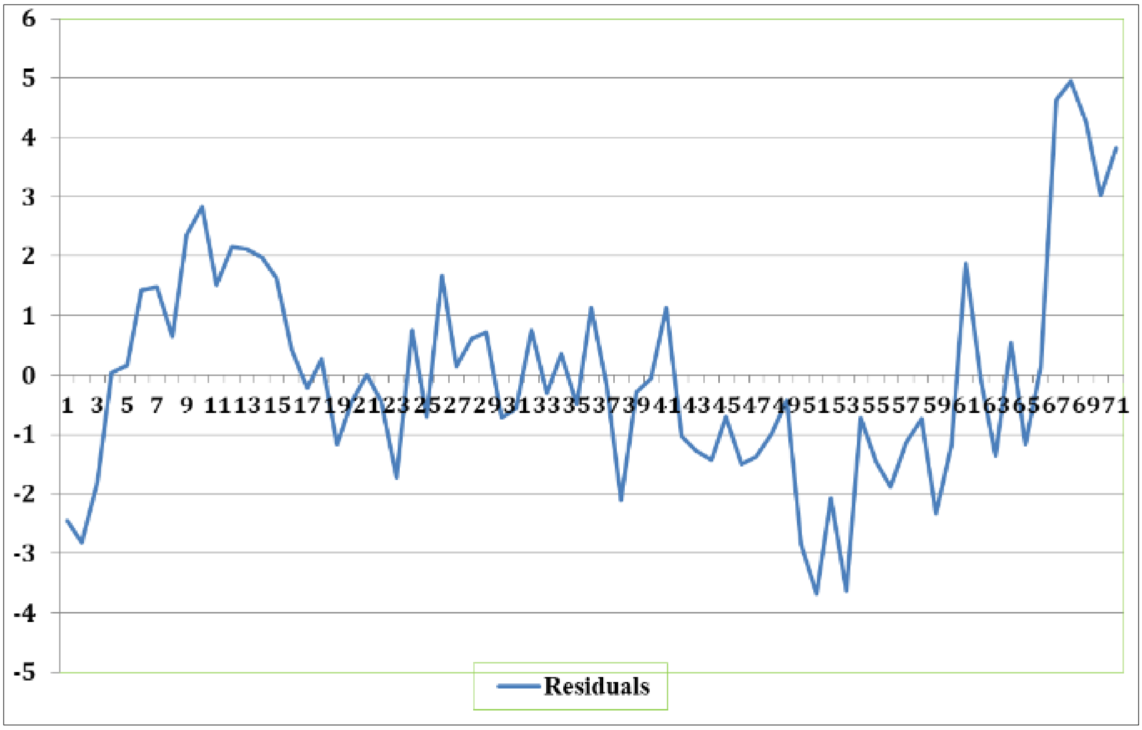

Fig 4: Plots of Residuals of Forecasted Pulses Area Data by Cubic Model

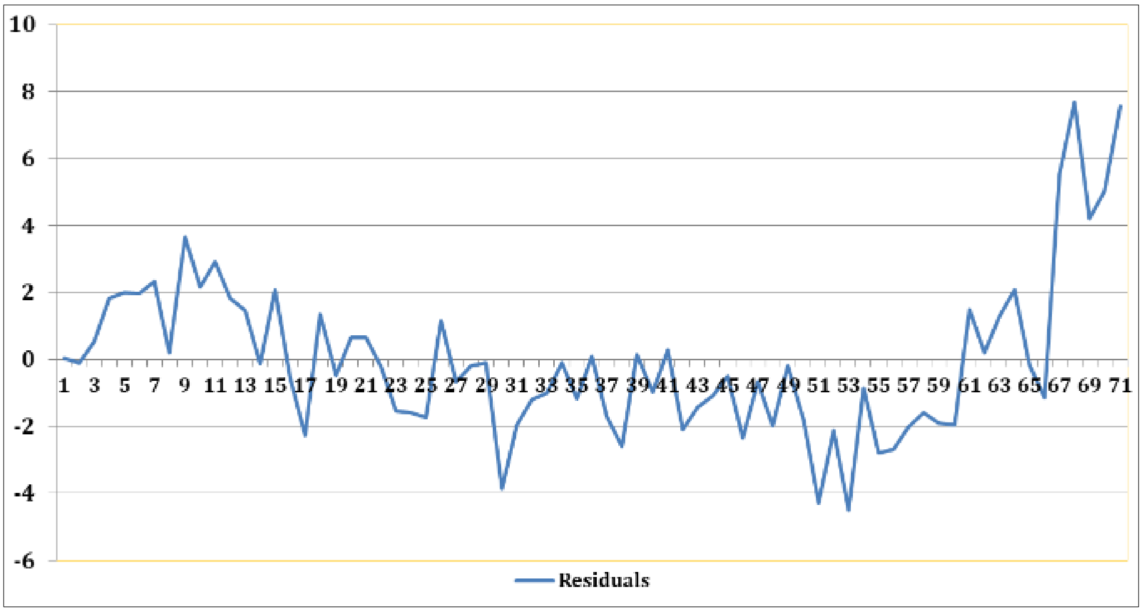

Fig 5: Plots of Residuals of Forecasted Pulses Productions Data by Cubic Model

The validation results of fitted models are described above. The calculated value of *F* is higher than the *F* table value which confirmed that Eq. (7) to (8) is significant at 95% confidence level. Thus, it was concluded that explanatory variables employed in this study for fitting the regression models for area and production of pulses are significant.

Forecasting pulses

The MAPE values calculated for cubic model developed for area and production were determined as 0.0027 and 2.1508 respectively (see Table 2, 3). These results suggested that cubic model which is selected as best forming model can give decent prediction of the area and production of pulses. The cubic models have MAPE values which indicated that both models developed for area and production of pulses are high accuracy prediction models (see Table 2 and 3). It is also clearly visible in from Fig.3 and 4 that the actual and predicted values for the area and production of pulses lying very close to each other.

Table 5: Forecasted values of Pulses Area and Production by Cubic Model

Year	Forecasted Area	Year	Forecasted Area
2022	29.91	2027	33.99
2023	30.64	2028	34.94
2024	31.41	2029	35.94
2025	32.23	2030	36.98
2026	33.09	2031	38.08

The area and production of pulses predicted using cubic model for next 10 years are showing in Table 5 and illustrated in Fig. 3 and 4 respectively. The results suggested that the area and production of pulses may reach up to 2.11 million hectare and 7.75 million tonnes in 2031.

Conclusions

In this study, regression modeling approach was used for predicting the area and production of pulses, which are defined for the prosperity of farmers in decision making and also helps the government in making of policy related to agriculture. The following are the main outcomes of the study:

- The study concluded that the cubic regression model outperformed all other models used in this study.
- The determination coefficients for the training data sets of area and production of pulses are 0.650 and 0.844 respectively, which indicated that fitted models are highly reliable.
- The F test showed a statistically significant result for the cubic model fitted for the area and production of pulses. The small values for MAPE and RMSE for the fitted models also concluded that best selected cubic model have high accuracy.

In future, the researchers must undergo application of other models like artificial intelligence, grey theory forecasting, hybrid approach, multiple linear regression and other econometric methods for modeling the area and production of pulses. The results derived from the future studies can be compared with the present study for evaluation.

Acknowledgement

The authors are expressing our gratitude towards the reviewer for the valid suggestions which enhanced the quality of the manuscript. The authors are also showing our gracefulness towards the PG and PhD research scholars, teaching and non-teaching staffs of Department of Agricultural Statistics, College of Agriculture, Kumarganj, Acharya Narendra Deva University of Agriculture and Technology, Ayodhya, Uttar Pradesh.

References

1. Directorate of Economics & Statistics, DAC&FW, New Delhi.https://eands.dacnet.nic.in/
2. https://pulses.dac.gov.in/areaproduction.aspx
3. Kumar S, Singh J. Forecasting of Pulses Production in India: Application of UCM Model. International journal in management and social science (Impact factor-5.276); c2016, 4.
4. Mandal S, Mondal K, Ghoshal S, Pal DBA, Acharya SK. Livelihood Implications of Pulse in an Operating Cropping System. Current Journal of Applied Science and Technology. 2020;39(29):91-99.
5. Narayan P, Kumar S. Constraints of growth in area production and productivity of pulses in India: An analytical approach to major pulses. Indian Journal of Agricultural Research; c2015, 49(2).
6. Yadav RS, Mehta V, Tiwari A. An application of time series ARIMA forecasting model for predicting nutria cereals area in India. The Pharma Innovation journal. 2022;SP-11(3):1260-1267.
7. Yadav RS, Mehta V, Gokul Krishnan KB. Regression models for forecasting Nutri cereals area and production in India. International Journal of Statistics and Applied Mathematics. 2023;8(3):48-53.
8. Reddy A, Reddy GP. Supply side constrains in production of pulses in India: Case study of lentils. Agricultural economics research review. 2010;23:129-136.
9. Sah U, Dixit GP, Kumar H, Ojha J, Katiyar M, Singh V, *et al*., Dynamics of pulse scenario in Bundelkhand region of Uttar Pradesh: A temporal analysis. Indian Journal of Extension Education. 2021;57(4):97-103.
10. Sharma MK, Sisodia BVS, Lal K. Growth and trends of pulse production in India. Journal of Food Legumes. 2013;26(1):86-92.
11. Supriya AB, Raghav YS, Devi M, Kumari P, Yadav S, Mishra P, *et al*., Modeling and forecasting of lentil production in India and its Instability. Journal of Animal &plant sciences. 2023;33(4):817-828.
12. Vani GK, Mishra P. Impact of irrigation on pulses production in India: A time-series study. Legume Research-An International Journal. 2019;42(6):806-811.